

This document relates to all cases.

¹ The reference to “Hogan Lovells Defendants” includes the following named defendants in this MDL action: Louisiana Health Service and Indemnity Company d/b/a/ Blue Cross and Blue Shield of Louisiana, BlueCross BlueShield of Tennessee, Inc., Blue Cross and Blue Shield of Vermont, BCBSM, Inc. d/b/a/ Blue Cross and Blue Shield of Minnesota, Wellmark, Inc. d/b/a/ Wellmark Blue Cross and Blue Shield of Iowa, Wellmark of South Dakota, Inc. d/b/a Wellmark Blue Cross and Blue Shield of South Dakota, Blue Cross and Blue Shield of Florida, Inc., Blue Cross and Blue Shield of Massachusetts, Inc., Blue Cross and Blue Shield of North Carolina, Inc., Blue Cross and Blue Shield of Rhode Island, BlueCross BlueShield of South Carolina, Inc., Cambia Health Solutions, Inc. (and subsidiaries named in the Complaint), Hawaii Medical Service Association d/b/a Blue Cross and Blue Shield of Hawaii, Horizon Healthcare Services, Inc. d/b/a Horizon Blue Cross and Blue Shield of New Jersey, and WellPoint, Inc. (and subsidiaries named in the Complaint).

requests made by the plaintiffs. In opposition to the motion to compel, the Hogan Lovells Defendants assert that they have agreed to make available to plaintiffs certain search metrics related to search terms over which there is disagreement, but not as to those terms on which the parties have reached agreement.

To set the stage for any casual reader of this Order (if such a person actually exists), the Provider Plaintiffs and the Subscriber Plaintiffs have propounded to the Hogan Lovells Defendants (and all others, as well) requests for production of documents that require searching massive amounts of electronically-stored information (“ESI”) maintained by the defendants in large and complex computer systems. To search for responsive documents among the mountains of ESI, the Hogan Lovells Defendants have decided to use a key-word or search-term search methodology, to which the plaintiffs have acquiesced.² This methodology requires

² An alternative form of relief sought by the motion is for the court to order the Hogan Lovells Defendants to abandon their key-word or search-term methodology and use instead a TAR (“technology-assisted review”) method of searching. The court will not do so. Not only have the parties previously agreed, even if reluctantly, to the search-term methodology, the choice of how to go about gathering the documents requested by a party rests, at least initially, with the producing party. The producing party usually is in the best position to address how to locate, access, and review the requested documents in the least expensive and most efficient manner. Unless the search methodologies of the producing party prove to be inadequate to the task of fairly identifying the responsive documents, the decision how to gather and produce the documents must be left to the producing party. To date, the search-term method has not been shown to be inadequate to the task. That, perhaps, remains to be seen.

crafting search terms³ that are used to query the computer system to retrieve documents fitting the search-term parameters. The words or abbreviations selected and the delimiters used implicate the inclusiveness of the search term. Depending upon the structure of the search term, it may result in “over-inclusiveness,” that is, identifying documents not relevant to the documents requested, or it may “under-inclusive,” failing to identify and capture plainly relevant documents.

The parties have spent months negotiating the structure of hundreds of search terms to try to find the proper structure for each term to identify the greatest number of responsive documents but without a significant number of irrelevant ones. What is at issue in this motion is the plaintiffs’ desire to have the Hogan Lovells Defendants disclose to them search metrics with respect to each search term being used to query the defendants’ computer databases to enlighten and assist them in the negotiations. These metrics include such things as “hit reports” (statistics showing the number of documents identified by each search term), false-negative sampling reports (sampling of responsive documents to determine the rate at which the search term *failed* to identify a responsive document), and “unique hit

³ The term “key word” does not do justice to the complexity of the terms created for these searches. The search terms can be long strings of words or abbreviations connected by “delimiters,” terms used to instruct the computer to search for the words or abbreviations in some relationship to each other, such as “within the same sentence” or “within 5 words” of each other. An example of a search term proposed by the plaintiffs in this case is “(“Sherman Act” OR “Clayton Act” OR “Section 1” or “Section 2” OR monopol* OR ((price* or rate*) 2N fix*) OR “Section 5” OR (unfair* 4N practice*) OR boycott OR conspire* OR conspirac* OR collusion OR collud*)100N ((blue* OR BCBS* or BCA or BSA or cross or shield or network).” See Hogan Lovells Defendants’ Ex. O, Doc. 570-15, ref no. 422.

reports” (reports of documents identified *only* by a particular search term and not by others). Plaintiffs argue they need these metrics as part of their negotiations with the Hogan Lovells Defendants on the proper structure of search terms because, without these testing and validation metrics, the plaintiffs remain in the dark about the efficacy of the search terms to hit an adequate number of responsive documents without being over-inclusive.⁴

After the filing of the plaintiffs’ motion to compel, the Hogan Lovells Defendants filed their response, asserting that, in fact, they have disclosed to plaintiffs many of the metrics sought, and they agreed “to continue to provide this or similar information to the Plaintiffs for additional search queries that the parties are still negotiating as their assessment continues.” (Doc. 570, p. 16). Also, they contend that they “have provided, and will continue to provide, hit counts to the Plaintiffs for queries that are still being negotiated.” (*Id.*). On March 17, 2016, it appears that Exhibit O to the Hogan Lovells Defendants’ opposition to the motion was emailed to counsel for the plaintiffs. Exhibit O is a combination of three charts showing (1) the “hit counts” for various search terms with modifications

⁴ Much as one might hear from former-athletes-turned-commentators on *Sports Center*, this controversy is a view into “the game within the game” of how complicated ESI discovery is conducted by the parties. The defendants refer to it as “discovery on discovery” (Doc. 570, pp. 18, 19), but plaintiffs make the point that they cannot monitor document production or agree to certain search terms proposed by the Hogan Lovells Defendants if they have no information about how the search terms actually play out when queried. But also, the court must keep in mind that this is not a true “discovery” dispute, but a dispute over how much information the defendants must share about their efforts to comply with plaintiffs’ requests to produce documents.

proposed by the Subscriber Plaintiffs and the Hogan Lovells Defendants (the 3-page chart),⁵ (2) the search terms (but not the hit counts) on which the parties have reached agreement (the 30-page chart),⁶ and (3) the permutations of search terms (but not hit counts for the various permutations) still being negotiated (the 20-page chart).⁷

While this agreement does not resolve entirely the plaintiffs' demand for

⁵ The court confesses that it does understand the meaning of the column on the 3-page chart titled "Delta Resp. Rate." The court assumed this to mean the rate of change (Δ , or *delta*) as a percentage in the number of responsive documents produced by a modified search term proposed by the Subscriber Plaintiffs compared to an earlier search term used by the Hogan Lovells Defendants. Either the court's misunderstanding of the columns on the chart or its limited math skills leaves it confused by this column. For example, there is a column entitled "Preliminary Hit Counts," in which it appears that the number of responsive documents identified by use of a search term defined by the Hogan Lovells Defendants is reported. Then, after another column reflects a modification of that term proposed by the Subscriber Plaintiffs, the seventh column appears to report the "Delta," or the number of additional documents produced by using the plaintiffs' proposed modification. The eighth column then records a "Delta Resp. Rate" in terms of, presumably, a percentage increase or degree in produced documents as a result of use of the modified search term proposed by the plaintiffs. This assumption cannot be right, however, because the court is unable to fathom the mathematical relationship between the "Preliminary Hit Count," the "Delta," and the "Delta Resp. Rate." For instance, taking Search Term 26, the "Preliminary Hit Count" for the number of documents retrieved using the Hogan Lovells Defendants' search term is reported as 225,900 documents. The "Delta" column reports that only 679 additional documents were retrieved by use of a modified search term proposed by the Subscriber Plaintiffs. The "Delta Resp. Rate" column, however, shows 10.91%. Although the court initially had assumed the "Delta Resp. Rate" was the percentage increase of documents the plaintiffs' search term "hit" compared to the number of documents "hit" by the defendants' search term, this plainly cannot be so. An increase of only 697 additional documents does not correspond to a 10.91% increase over the original hit count of 225,900. Thus, the court is left befuddled by the meaning of the "Delta Resp. Rate" column on the 3-page chart.

⁶ The plaintiffs' motion, filed March 9, 2016, argues that the parties have reached agreement on only about one-third of the 750+ search terms being discussed. However, the parties held another meet and confer after that date and, according to the Hogan Lovells Defendants' Exhibit O, agreement now has been reached on all but about 200 search terms. Plaintiffs have not indicated this was not correct as of the date of the March 23 discovery conference.

⁷ The Hogan Lovells Defendants refer to these three charts, respectively, as the "Modified" terms, the "Agreed" terms, and the "Pending" terms.

validation metrics, it provides a starting point for discussion. Providing “hit counts” produced by various permutations of search terms allows the plaintiffs to get a rough idea of the changes such modifications in search terms have on the burden and breadth involved in responding to them. Based on the agreement of the Hogan Lovells Defendants, the motion to compel is GRANTED and these defendants are DIRECTED to produce hit counts similar to those reflected in the 3-page “Modified” chart in Exhibit O for examination by the plaintiffs, but only with respect to those search terms on which no agreement has been reached. Defendants need not produce hit counts or reports as to agreed search terms.

Hit counts and reports do not provide insight into the under-inclusiveness of a search term, a term that fails to identify plainly responsive documents. The Hogan Lovells Defendants acknowledge that they are performing validation tests on the search terms *they* designed by sampling responsive documents to determine the rate of “false negatives” produced by each search term. A false negative occurs when a search term fails to capture a responsive document.⁸ A search term with a high rate of false negatives may be defectively under-inclusive. Although ESI discovery does not aspire to perfection, it must be adequate to comply fairly with the needs of the parties to find and retrieve evidence relevant and proportional to

⁸ A search term may fail to capture a responsive document for any number of reasons. The delimiters may be too narrow, or the wrong root terms are used, or there are an insufficient number of synonyms or expanders.

the case. In order to assure the plaintiffs that the search terms being negotiated are not defectively *under*-inclusive, it is ORDERED that the validation sampling of the *defendants'* search terms revealing a search term with a false-negative rate greater than ten percent (10%) must be provided to the plaintiffs. Rates at or less than ten percent need not be provided, nor are the Hogan Lovells Defendants required to conduct validation sampling on the search terms proposed by the plaintiffs.

Finally, the motion to compel is DENIED as to any requirement that “unique hit” counts or reports be supplied. A “unique hit” is one in which a particular search term identifies a responsive document or set of documents that no other search term captured. To assess whether a hit is “unique” requires a comparison of the captured document(s) to all other documents identified by every other search term to determine whether any other search-term query found the allegedly unique document(s).⁹ If the document(s) was retrieved by another search term or terms, the search term under consideration was not “unique,” that is, it was not the only search term that found the document(s).

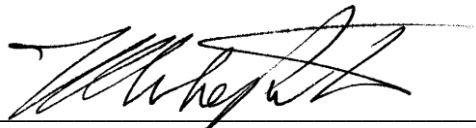
This is truly the needle in the haystack. Despite the hundreds of pitchforks

⁹ Of course, because the defendants will be supplying “hit” counts to the plaintiffs as to both their original and plaintiffs’ modified search terms, plaintiffs will be able to see those search terms that retrieve a small number of documents, perhaps suggesting uniqueness. The court understands that this raw number itself will not establish whether the documents so retrieved are “unique” in the sense that no other search-term query discovered them. Nonetheless, the burden and expense of making the comparison to the results of all other search-term queries necessary to determine uniqueness is not proportional to this particular issue, which, as already mentioned, is not a *true* discovery issue.

combing through the haystack, the needle was missed until just the right pitchfork with just the right balance and tines came along. The idea of Rule 26(b)(1) proportionality in ESI does not require (perhaps even does not allow) this level of perfection. Discovery must be fair, adequate, and proportional to the needs of the case, and while these MDL cases are very complex, expensive, and important, proportionality must limit the depth of effort and expense used to find the needle. Courts have long noted that discovery is not a fishing expedition, casting about here and there just to see what might turn up. Even more so today, in the age of overwhelming ESI, discovery must be more focused and guided. Given these legal and practical limits, and the recognition that ESI discovery is not perfect, the inadvertent loss of a “unique” document¹⁰ somewhere in the process is the practical price to be paid for reasonable and proportional discovery.

Except as provided herein, the motion to compel is DENIED in all other respects.

DONE this 21st day of April, 2016.

A handwritten signature in black ink, appearing to read 'T. Michael Putnam', written over a horizontal line.

T. MICHAEL PUTNAM
UNITED STATES MAGISTRATE JUDGE

¹⁰ By “unique document” the court means a document or documents that might have been found if only another, slightly different search term had been used.